

(19) 世界知的所有権機関
国際事務局



(43) 国際公開日
2005 年 9 月 9 日 (09.09.2005)

PCT

(10) 国際公開番号
WO 2005/083890 A1

(51) 国際特許分類: H03M 7/30, G06F 17/30
(21) 国際出願番号: PCT/JP2004/002252
(22) 国際出願日: 2004 年 2 月 26 日 (26.02.2004)
(25) 国際出願の言語: 日本語
(26) 国際公開の言語: 日本語
(71) 出願人 (米国を除く全ての指定国について): 三菱電機株式会社 (MITSUBISHI DENKI KABUSHIKI KAISHA) [JP/JP]; 〒1008310 東京都千代田区丸の内二丁目 2 番 3 号 Tokyo (JP).
(72) 発明者; および
(75) 発明者/出願人 (米国についてのみ): 高山 茂伸

(TAKAYAMA, Shigenobu) [JP/JP]; 〒1008310 東京都千代田区丸の内二丁目 2 番 3 号 三菱電機株式会社内 Tokyo (JP). 東 辰輔 (AZUMA, Shinsuke) [JP/JP]; 〒1008310 東京都千代田区丸の内二丁目 2 番 3 号 三菱電機株式会社内 Tokyo (JP). 佐藤 重雄 (SATO, Shigeo) [JP/JP]; 〒1008310 東京都千代田区丸の内二丁目 2 番 3 号 三菱電機株式会社内 Tokyo (JP).

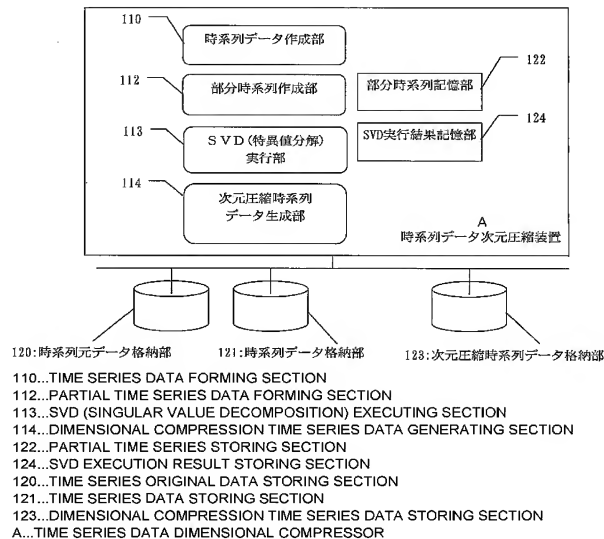
(74) 代理人: 溝井 章司, 外 (MIZOI, Shoji et al.); 〒2470056 神奈川県鎌倉市大船二丁目 1 7 番 1 0 号 N T A 大船ビル 3 階 溝井国際特許事務所 Kanagawa (JP).

(81) 指定国 (表示のない限り、全ての種類の国内保護が可能): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, HR, HU,

[続葉有]

(54) Title: TIME SERIES DATA DIMENSIONAL COMPRESSOR

(54) 発明の名称: 時系列データ次元圧縮装置



(57) Abstract: A time series data dimensional compressor performing dimensional compression for improving a retrieval efficiency of time series data without losing the features of data. The time series data is compressed to a predetermined dimension such that a larger number of pieces of information can be extracted. A partial time series forming section (112) forms a partial time series divided into a specified segment width for a plurality of elements of time series data generated at a time series data forming section (110). A singular value decomposition executing section (113) performs singular value decomposition on all partial time series, and a dimensional compression time series data generating section (114) generates dimensional compression time series data using a high-order component of singular value decomposition as a representative value of partial time series.

(57) 要約: 時系列データ次元圧縮装置に係り、データの特徴を失わずに、時系列データの検索の効率をよくする次元圧縮を行うことを課題とする。また、決まった次元に圧縮し、そのなかにより多くの情報を抽出できるようにする。部分時系列作成部 112 は、時系列データ作成部 110 で生成した複数の時系列データについて、指定のセグメント幅に分割した部分時系列を作成する。特異値分解実行部 113 は、

[続葉有]



WO 2005/083890 A1



ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NA, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.

CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IT, LU, MC, NL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

添付公開書類:

— 国際調査報告書

(84) 指定国 (表示のない限り、全ての種類の広域保護が可能): ARIPO (BW, GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), ユーラシア (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), ヨーロッパ (AT, BE, BG, CH,

2文字コード及び他の略語については、定期発行される各PCTガゼットの巻頭に掲載されている「コードと略語のガイダンスノート」を参照。

明 細 書

時系列データ次元圧縮装置

5 技術分野

本発明は、時系列データの検索をより効率よく行うために、データの特徴を失わずに次元圧縮することを目的とする。また、圧縮効率をよくすることを目的とするのではなく、決まった次元に圧縮し、そのなかにより多くの情報を抽出することを目的とする。

10

背景技術

時系列データの次元を減らす従来技術としては、例えば
E. Keogh, K. Chakrabarti, M. Pazzani
, Mehrotra

15 “Dimensionality Reduction for Fast Similarity Search in Large Time Series Databases.” Journal of Knowledge and Information Systems, 2000

20 に示されているPAA (Piecewise Aggregate Approximation) がある。

PAAでは、時系列データをセグメントに分割し、セグメントの平均値を各セグメントの代表値とすることで、時系列データを圧縮するものである。

25 平均値の計算はフーリエ変換や特異値分解に比べて簡単に実行でき、より高速に次元圧縮時系列データを生成することができる。

時系列データの次元を減らす他の従来技術としては、例えば、
F. Korn, H. V. Jagadish, C. Faloutsos
"Efficiently Supporting Ad Hoc Q
ueries in Large Datasets of Time
5 Sequences"
Proceedings of SIGMOD '97 pp 289-
300

に示されている特異値分解による方法もある。特異値分解による方法では、特異値分解後の全ての成分を用いるのではなく、主要な特異値（特
10 異値の大きいもの）のみを採用することで時系列データを圧縮するものである。

特異値分解により次元を圧縮すると、データの形状を他の方法よりうまく抽出できるため、検索効率がよいという利点がある。

また、画像データの次元を減らす従来技術としては、例えば特開昭 6
15 1-285870「変換符号化方式」がある。画像データをブロックに分割してブロックごとに圧縮する。分割したブロックの圧縮では、DCT
(Discrete Cosine Transform) と行列の縦・横の傾斜角度をあらわす変換との組合せを用いる。

このように2つの変換を組み合わせることで、ブロックごとにその特徴
20 徴を抽出して最適な変換を選択することでより高い圧縮率を実現できる。

PAAは、各セグメントの平均値を代表値とすることで、より高速に次元圧縮ができるが、時系列データの検索時または類似検索時には次のような問題がある。時系列データの検索手順は、まず圧縮空間で解の候補
25 補を探し、各解の候補に対して実空間で最終的な解を検索する。したがって、圧縮空間では解の候補として選択されたが実空間で真の解となら

ないものが多いと検索効率が悪くなる。P A Aの場合は平均値を各セグメントの代表値とするために、時系列の形状が消されるため圧縮後の情報が乏しくなり、上記で言う検索効率が悪いという問題がある。例えば、平らな時系列も右上がりの時系列も右下がりの時系列も平均値が同じ
5 になると、圧縮後の値は同じになってしまう。

S V Dは、特異値分解によりデータの形状をうまく抽出でき、上記で言う検索効率は高いものの、データ量が多くなると特異値分解にかかる時間が多くなり現実的な時間では特異値分解ができないという問題がある。

10 特開昭61-285870「変換符号化方式」は、圧縮率を向上することが目的であるが、時系列データの検索に用いる場合には以下の問題がある。時系列データの検索はまず解の候補を圧縮空間で探すために、全てのセグメント（ブロック）を同じ圧縮率で圧縮する必要があるが、上述の方式では各ブロックごとに異なる圧縮率となる。

15

発明の開示

本発明に係る時系列データ次元圧縮装置は、

以下の要素を有することを特徴とする

20 (1) 時間軸に沿って一定間隔で測定された一連のデータである時系列元データに対して、時間軸上で所定間隔ずつ開始位置をずらして、指定の長さの時系列データを複数作成する時系列データ作成部

(2) 複数の上記時系列データのそれぞれについて、指定のセグメント幅に分割した部分時系列を作成する部分時系列作成部

25 (3) 上記分割した全ての部分時系列を対象として特異値分解を行う特異値分解実行部

(4) 指定された数の上位の特異値分解の成分を、上記各セグメント幅

に分解した部分時系列の代表値として、次元圧縮時系列データを生成する次元圧縮時系列データ生成部。

図面の簡単な説明

- 5 図 1 は、この発明の実施の形態 1 を示す構成図である。
- 図 2 は、時系列元データ 1 5 0 を表したグラフである。
- 図 3 は、時系列データ 1 5 1 の作成方法の図である。
- 図 4 は、時系列データ 1 5 1 の作成のためのフローチャートである。
- 図 5 は、時系列データ 1 5 1 のグラフである。
- 10 図 6 は時系列データをセグメントに分割した様子を示したものである。
- 図 7 は、開始位置 $t = k + 2N$ の部分時系列 1 5 2 を示したものである。
- 図 8 は、 k から始まる時系列 1 5 1 と $k + 2N$ から始める時系列 2 5
- 15 1 を示したものである。
- 図 9 は、SVD 実行結果記憶部の内容で特異値分解の結果を示す図である。
- 図 10 は、代表値をプロットした次元圧縮時系列データ 1 5 3 のグラフの例である。
- 20 図 11 は、圧縮データ作成のフローチャートである。
- 図 12 は、この発明の実施の形態 2 を示す構成図である。
- 図 13 は、実施の形態 2 のフローチャートである。
- 図 14 は、セグメント幅を 16 にした場合と 32 にした場合の模式図である。
- 25 図 15 は、SVD の結果の第 2 成分まで用いる場合の SVD 実行結果である。

図 1 6 は、この発明の実施の形態 3 を示す構成図である。

図 1 7 は、平均値計算実行部 1 8 2 の計算結果を模式的に新たしたものである。

図 1 8 は、中間時系列のグラフである。

5 図 1 9 は、圧縮後の次元を 8 次元にする場合の S V D 実行結果である。

図 2 0 は、ハードウェア構成図である。

発明を実施するための最良の形態

10 実施の形態 1 .

図 1 は、この発明の実施例を示す構成図である。図において、1 2 0 は 2 次記憶装置もしくは 1 次記憶装置で、時系列元データ 1 5 0 を格納した時系列元データ格納部である。時系列データ作成部 1 1 0 は、時系列元データ格納部 1 2 0 から時系列元データ 1 5 0 を読み込み、時系列データ 1 5 1 を作成する。1 2 1 は 2 次記憶装置もしくは 1 次記憶装置で 1 1 0 によって生成された複数の時系列データ 1 5 1 を格納する時系列データ格納部である。部分時系列作成部 1 1 2 は、時系列データ格納部 1 2 1 から時系列データ 1 5 1 を順次読み込み、部分時系列 1 5 2 を作成し、部分時系列記憶部 1 2 2 に格納する。部分時系列記憶部 1 2 2 は、
20 、 1 次記憶装置もしくは 2 次記憶装置である。S V D 実行部 1 1 3 は、部分時系列記憶部 1 2 2 から部分時系列 1 5 2 を読み込み特異値分解を実行し、S V D 実行結果記憶部 1 2 4 に格納する。S V D 実行結果記憶部 1 2 4 は、1 次記憶装置もしくは 2 次記憶装置である。次元圧縮時系列データ生成部 1 1 4 は、S V D 実行結果記憶部 1 2 4 から S V D の結果
25 を読み込み、次元圧縮時系列データ 1 5 3 を作成し、次元圧縮時系列データ格納部 1 2 3 に格納する。次元圧縮時系列データ格納部 1 2 3 は、2

次記憶装置もしくは1次記憶装置である。

図2は、時系列元データ150を表したグラフである。x軸が時刻 t 、y軸が時系列の値である。時刻 t の取りうる値は1から m の間の自然数であり、 m 個のデータポイントがある。最初のデータポイントは $t=$
5 1であらわし、最後のデータポイントは $t=m$ とする。また、データポイントの数を長さと呼び、この場合の長さは m （時系列長 160）である。

図3は、時系列データ151の作成方法の図である。時系列データ作成部110は、時系列元データ格納部120から時系列元データ150
10 を読み込み、長さ n の時系列を時系列元データの開始ポイントの時刻 t を一つづつずらすことで、 $m-n+1$ 個作成する。ここでは、長さ n はあらかじめ決めておくこととする。 $t=1$ から始める時系列が時系列1、 $t=2$ から始める時系列が時系列2、そして $t=m-n+1$ から始める時系列が時系列 $m-n+1$ である。 $t=m-n+1$ から始める時系列
15 の終点は $t=m$ となり、これ以降の t の値から始めると長さが n 未満となる。

さらに、部分時系列作成のことを考慮し、長さが n 未満で N 以上の時系列をその後に $n-N$ 個追加する。これを補填時系列とよぶ。開始時刻
20 t の値が、 $m-n+2$ から $m-N+1$ までの時系列であり、終点の t の値は全て m である。

$m-n+2$ から始まる時系列は、長さが $n-1$ であり、
 $m-n+3$ から始まる時系列は、長さが $n-2$ であり、
 $m-N+1$ から始まる時系列は、長さが N である。

図4は、時系列データ151の作成のためのフローチャートである。
25 S301にて、時系列元データにて時系列データの開始時刻 $t=1$ にセットする。S302においては、時系列長 $=n$ にセットする。S303

にて、時系列元データを読み込む。S 3 0 4 では、時系列開始時刻と時系列長から時系列終了ポイントを計算し、 m 以下であるか調べる。時系列終了ポイントが m 以下であれば時系列データを作成できるので、S 3 0 5 にすすむ。S 3 0 5 にて時系列開始時刻と時系列長をもとに、時系列元データから時系列データを作成する。S 3 0 6 では次の時系列データを作成するために開始ポイント t をインクリメントして、再びS 3 0 3 に戻る。S 3 0 4 にて、時系列の終了ポイントが m を超えている場合は、時系列長 n の時系列データはこれ以上作成できないので、補填時系列データを作成するためにS 3 0 8 にすすむ。S 3 0 8 では、時系列長をデクリメントする。S 3 0 9 では、デクリメントして時系列長が N 以上であるか調べる。 N 以上であれば、S 3 1 0 にすすみ補填時系列データを作成する。S 3 1 1 では、開始ポイントをインクリメントして次の補填時系列データを作成するための準備をして、S 3 0 7 にすすむ。S 3 0 7 では時系列元データを読み込む。次に再びS 3 0 8 にすすむ。S 3 0 9 にて、時系列長が N より小さくなったら時系列データ作成は終了する。

図 5 は、時系列データ 1 5 1 のグラフである。開始位置は k 、終了位置が $k + n - 1$ で n 個のデータポイントで構成される。時系列データ 1 5 1 は長さが n (検索時系列長 1 6 1) の時系列である。

図 6 は時系列データをセグメントに分割した様子を示したものである。各時系列データ 1 5 1 を長さ N (セグメント幅 1 6 2) のセグメントに分割する。一つの時系列データ 1 5 1 は n/N 個のセグメントに分割される。それぞれの長さ N のセグメントを部分時系列 1 5 2 とする。

図 7 は、開始位置 $t = k + 2 N$ の部分時系列 1 5 2 をあらわす。部分時系列 1 5 2 は N 個のデータポイントから構成され、長さは N である。

部分時系列作成部 1 1 2 は、それぞれの時系列データ 1 5 1 の最初の
 N個のデータのみを選択することで、部分時系列データ 1 5 2 を作成す
 る。全ての時系列に対して実行することで開始位置が $t = 1$ から $t = m$
 $- n + 1$ の長さ N の部分時系列を作成できる。さらに、時系列データ作
 成部 1 1 0 が作成した補填時系列についても最初の N 個のデータを読み
 込みそれも補填部分時系列データとして作成し、あわせて部分時系列記
 憶部 1 2 2 に格納する。ここでセグメント幅 N はあらかじめ決めておく
 こととする。上記のようにすることで、時系列元データに対して、開始
 位置 $t = 1$ から $t = m - N + 1$ までの長さ N の全ての部分時系列を作成
 できる。

全ての時系列データがもともとは一つの時系列元データ 1 5 0 であっ
 たことから、各時系列データのいずれのセグメントも上記の部分時系列
 のいずれかと一致する。

図 8 に示すように、k から始まる時系列 1 5 1 の 3 番目のセグメント
 の部分時系列は、 $k + 2N$ から始める時系列 2 5 1 の最初のセグメント
 と同じである。つまり、時系列 2 5 1 から作成した部分時系列と一致す
 る。

SVD 実行部 1 1 3 は、部分時系列作成部 1 1 2 から部分時系列デー
 タ 1 5 2 を読み込み、行数 $m - N + 1$ 、列数 N の行列として特異値分解を
 実行する。

特異値分解とは、任意の $m \times n$ 行列 Y を、以下のように U, S, V の 3
 つの行列の積で表わすもので一般に良く知られている式である。

$$Y = USV^T = [u_1, u_2, \dots, u_r] \begin{bmatrix} s_1 & 0 & \dots & 0 \\ 0 & s_2 & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \dots & 0 & s_r \end{bmatrix} \begin{bmatrix} v_1^T \\ v_2^T \\ \vdots \\ v_r^T \end{bmatrix} = s_1 u_1 v_1^T + s_2 u_2 v_2^T + \dots + s_r u_r v_r^T$$

ただし、 $r = \text{rank}(Y)$,

s_1, s_2, \dots, s_r は $Y^T Y$ の正の固有値の平方根（特異値という）で、 $s_1 \geq s_2 \geq \dots \geq s_r$,

5 v_1, v_2, \dots, v_r は n 次のベクトルで、 $Y^T Y$ の固有値 $s_1^2, s_2^2, \dots, s_r^2$ に対応する固有ベクトル。

v_1, v_2, \dots, v_r は大きさが 1 で互いに直交する。

u_1, u_2, \dots, u_r は m 次のベクトルで、

$u_j = \frac{1}{s_j} Y v_j (j=1, 2, \dots, r)$ で定義される。

10 U は u_1, u_2, \dots, u_r を列にもつ $m \times r$ 行列

,

V は v_1, v_2, \dots, v_r を列にもつ $n \times r$ 行列

,

15 S は s_1, s_2, \dots, s_r を対角要素にもつ r 次の対角行列。

図 9 は、SVD 実行結果記憶部の内容で特異値分解の結果を示す。特異値分解により、各行の代表値として上記 $u_1 s_1$ を抽出する。

すなわち特異値分解の対象となる行数 $m - N + 1$ 列数 N の行列に対して、行方向で見た場合に、 r 番目の行は u_1 ベクトルの r 番目の成分と s_1 の積を代表値とする。

20 r 番目の行は、開始位置の $t = r$ で始まる部分時系列であり、その代表値が u_1 ベクトルの r 番目の成分と s_1 の積である。SVD 実行部は、全てのセグメント（全ての部分時系列）の代表値を作成する。

25 次に、次元圧縮時系列データ生成部は 114 は、上記の特異値分解の第一成分を各セグメントの代表値として、次元圧縮時系列データを生

成する。開始位置 $t = k$ で始まる部分時系列 1 5 1 は、以下の部分時系列で構成される。

開始位置 $t = k, k + N, k + 2N, \dots$ 。

したがって、その次元圧縮時系列データは、最初の代表値は u_1 ベクトルの k 番目の成分と s_1 の積であり、次の代表値は u_1 ベクトルの $k + N$ 番目の成分と s_1 の積である。

図 1 0 は、上記の代表値をプロットした、次元圧縮時系列データ 1 5 3 のグラフの例である。

次元圧縮時系列データ 1 5 3 は n/N 個のポイントで構成される。時系列データ 1 5 1 をセグメント分割したそれぞれの部分時系列データにたいして、SVD を実行してその第一成分をプロットしたものである。

図 1 1 は、圧縮データ作成のフローチャートである。時系列データ作成部 1 1 0 は、時系列元データ格納部 1 2 0 から時系列元データ 1 5 0 を読み込み、時系列データ 1 5 1 を作成し、時系列データ格納部 1 2 1 に格納する。次に、部分時系列作成部 1 1 2 は、時系列データ格納部 1 2 1 から時系列データ 1 5 1 を順次読み込み、部分時系列 1 5 2 を作成し、部分時系列記憶部 1 2 2 に格納する。次に、SVD 実行部 1 1 3 は、部分時系列記憶部 1 2 2 から部分時系列を読み込み特異値分解を実行し、SVD 実行結果記憶部 1 2 4 に格納する。次元圧縮時系列データ生成部 1 1 4 は、SVD 実行結果記憶部 1 2 4 のデータを用いて次元圧縮時系列データ 1 5 3 を作成し、次元圧縮時系列データ格納部 1 2 3 に格納する。

上述のように時間軸に沿って一定間隔で測定された一連のデータに対して、指定の長さの時系列データを時間軸上で開始位置をずらして複数作成する手段と、上記指定の長さの各時系列データを指定のセグメント幅に分割した部分時系列を作成する手段と、上記分割した全ての部分時

系列を対象として特異値分解を行う手段と、指定された数の上位の特異値分解の成分（この場合は第一成分までとした）を上記各セグメント幅に分解した部分時系列の代表値とする手段と、上記代表値を組み合わせることで上記指定の長さの時系列データの次元を圧縮する手段とを備えたことを特徴とする時系列データ次元圧縮方式について説明した。

実施の形態 2.

図 1 2 は、この発明の実施例を示す構成図である。1 1 0、1 1 2、1 1 3、1 1 4、1 2 0、1 2 1、1 2 2、1 2 3、1 2 4 は図 1 と同様である。データ解析部 1 1 7 は、時系列データ格納部 1 2 1 から時系列データ 1 5 1 を読み込み解析し、セグメント幅と特異値分解の結果のどの成分までを有効とするかを定めるものである。

図 1 3 は実施の形態 2 のフローチャートである。時系列データ作成部 1 1 0 は、時系列元データ格納部 1 2 0 から時系列元データ 1 5 0 を読み込み、時系列データ 1 5 1 を作成し、時系列データ格納部 1 2 1 に格納する。

次に、データ解析部 1 1 7 は時系列データ格納部 1 2 1 から時系列データを読み込み解析を行う。解析の結果、検索時にヒット率が最も高くなるように、セグメント幅と特異値分解の結果のどの成分までを有効とするかを定める。この場合は、第 2 成分までを使用することとする。

部分時系列作成部 1 1 2 は、時系列データ格納部 1 2 1 から時系列データ 1 5 1 を順次読み込み、部分時系列 1 5 2 を作成し、部分時系列記憶部 1 2 2 に格納する。この際に、部分時系列のセグメント幅はデータ解析部 1 1 7 が決めた値とする。次に、SVD 実行部 1 1 3 は、部分時系列記憶部 1 2 2 から部分時系列を読み込み特異値分解を実行する。実行した結果、SVD の結果のどの成分までを採用するかデータ解析部 1 1 7

が決めた値までをSVD実行結果記憶部に記憶する。この場合は第2成分までをSVD実行結果記憶部に格納する。次元圧縮時系列データ生成部114は、SVD実行結果記憶部の内容を用いて次元圧縮時系列データ153を作成し、次元圧縮時系列データ格納部123に格納する。

- 5 図14は、セグメント幅を16にした場合と32にした場合の模式図である。セグメント幅を16にして、SVDの第1成分まで使用した場合の圧縮後の次元は、以下より求まる。

セグメント数 $128 \div 16 = 8$ 、セグメント代表値 = 1

セグメント数×セグメント代表値 = 8

- 10 すなわち8次元に圧縮できる。

一方、セグメント幅を32にして、SVDの第2成分まで使用した場合の圧縮後の次元は、以下より求まる。

セグメント数 $128 \div 32 = 4$ 、セグメント代表値 = 2

セグメント数×セグメント代表値 = 8

- 15 すなわち8次元に圧縮できる。

圧縮後の次元を同じにする場合でも、セグメント幅とセグメント代表値の取り方にはいくつかの選択肢があり、そのなかでもっともヒット率が高くなるようにセグメント幅とセグメント代表値の数をきめるのがデータ解析部117の機能である。

- 20 図15は、SVDの結果の第2成分まで用いる場合のSVD実行結果記憶部の内容である。セグメントの第1成分が全セグメントでほぼ同じ値の場合には、セグメント幅を長くしてSVDの第2成分までを利用することでより詳細に時系列データの特徴を抽出でき、検索時のヒット率が向上する。

- 25 時系列データを解析して、上記時系列データを分割するセグメント幅および特異値分解のどの成分までを利用して上記部分時系列の代表値と

するかを判断する手段をそなえた請求項 1 記載の時系列データ次元圧縮方式について説明した。

5 以上のように、この発明によれば、セグメント分割して SVD を実行することで、全データと比較した各セグメントの特徴を抽出できるために、検索効率の高い圧縮データを作成できる。また、単純に SVD を実行する場合に比べると、対象となる行列の行数は同じであるが列数が N/n になるためにより高速に実行できる。

実施の形態 3.

10 図 16 は、この発明の実施例を示す構成図である。図において、110、114、120、121、123、124 は図 1 と同様である。中間次元決定部 181 は、平均値を計算するための幅を決定する。平均値計算実行部 182 は中間次元決定部の指定した平均値幅で時系列データの平均値を計算し、結果を平均値計算結果記憶部 191 に格納する。中
15 間時系列作成部 183 は、平均値幅の代表値をその平均値とすることで中間時系列 155 を作成して、中間時系列記憶部 192 に格納する。SVD 実行部 113 は、中間時系列記憶部 192 に対して、特異値分解を実行する。

20 中間次元決定部 181 は、時系列元データを読み込み解析して、中間次元 p および平均をとるセグメント幅を決定する。平均をとる幅は、時系列データが単調に増加または減少する範囲内とする。

図 17 は、平均値計算実行部 182 の計算結果を模式的に新たしたものである。時系列 151 の長さを n とし、中間次元を p とすると、平均をとるセグメント幅は n/p となる。例えば、時系列長が 128 で中間
25 次元が 32 の場合は、平均をとるセグメント幅は $128/32=4$ となる。平均値計算実行部 182 は、時系列元データ 150 に対して開始時

刻 t を一つづつずらしながら、データポイントごとに平均値を計算して結果を平均値計算結果記憶部 1 9 1 に格納する。

図 1 8 は、中間時系列のグラフである。中間時系列作成部 1 8 3 は、各時系列 1 5 1 を平均をとるセグメント幅に分解し、それぞれの代表値
5 を平均値計算結果記憶部 1 9 1 の内容から取り出して、中間時系列 1 5 5 を作成して、中間時系列記憶部 1 9 2 に格納する。

図 1 9 は、圧縮後の次元を 8 次元にする場合の S V D 実行結果記憶部 1 2 4 の内容である。S V D 実行部 1 1 3 は、中間時系列記憶部 1 9 2 から中間時系列 1 5 5 を読み込み、行数 $m - n + 1$ 、列数 p の行列として
10 特異値分解を実行し、結果を S V D 実行結果記憶部 1 2 4 に格納する。圧縮後の次元を 8 次元にするために、第 8 成分の値までを格納している。

次に、次元圧縮時系列データ生成部は 1 1 4 は、上記の特異値分解の第 8 成分までを用いて次元圧縮時系列データを生成する。すなわち、
15 各時系列 1 5 1 は以下の 8 つのデータで近似的に表し、次元圧縮時系列データを生成する。

$$(s_1 u_1, s_2 u_2, s_3 u_3, s_4 u_4, s_5 u_5, s_6 u_6, s_7 u_7, s_8 u_8)$$

指定の長さの複数の時系列データに対して、平均をとるセグメント幅
20 を決める手段と、上記の各時系列に対して平均をとるセグメント幅ごとに平均値を計算する手段と、平均値をセグメント代表値とした中間時系列を作成する手段と、それぞれの中間時系列を対象として特異値分解を行う手段と、指定された数の上位の特異値分解の成分を上記中間時系列の圧縮データとする手段とを備えたことを特徴とする時系列データ次元
25 圧縮方式について説明した。

以上のように、この発明によれば、時系列データが単調に変化する幅

で平均値をとることでデータの特質を失うことなくデータ量を削減できる。さらに、削減したデータ量に対してSVDを実行することで高速に特異値分解を実行でき、またデータの特徴も抽出することができる。

5 時系列データ次元圧縮装置は、コンピュータであり、各要素はプログラムにより処理を実行することができる。また、プログラムを記憶媒体に記憶させ、記憶媒体からコンピュータに読み取られるようにすることができる。

10 図20は、時系列データ次元圧縮装置のハードウェア構成例を示す図である。この例では、バスに、演算装置2001、メモリ2002、ハードディスク2003、表示装置2004が接続されている。プログラムは、例えばハードディスク2003に記憶されており、メモリ2002にロードされた状態で、順次演算装置2001に読み込まれ処理を行う。

15 産業上の利用可能性

データの特徴を失わずに、時系列データの検索の効率をよくする次元圧縮を行うことができる。また、決まった次元に圧縮し、そのなかにより多くの情報を抽出することができる。

請求の範囲

1. 以下の要素を有することを特徴とする時系列データ次元圧縮装置

- 5 (1) 時間軸に沿って一定間隔で測定された一連のデータである時系列元データに対して、時間軸上で所定間隔ずつ開始位置をずらして、指定の長さの時系列データを複数作成する時系列データ作成部
- (2) 複数の上記時系列データのそれぞれについて、指定のセグメント幅に分割した部分時系列を作成する部分時系列作成部
- 10 (3) 上記分割した全ての部分時系列を対象として特異値分解を行う特異値分解実行部
- (4) 指定された数の上位の特異値分解の成分を、上記各セグメント幅に分解した部分時系列の代表値として、次元圧縮時系列データを生成する次元圧縮時系列データ生成部。

- 15 2. 時系列データ次元圧縮装置は、上記代表値を組み合わせることで上記指定の長さの時系列データの次元を圧縮することを特徴とする請求項 1 記載の時系列データ次元圧縮装置。

3. 前記時系列データを解析して、当該時系列データを分割するセグメント幅および特異値分解について、どの成分までを利用して
- 20 上記部分時系列の代表値とするかを判断するデータ解析部を備えることを特徴とする請求項 1 記載の時系列データ次元圧縮装置。

4. 以下の要素を有することを特徴とする時系列データ次元圧縮装置

- (1) 時間軸に沿って一定間隔で測定された一連のデータである時系列
- 25 元データに対して、時間軸上で所定間隔ずつ開始位置をずらして、指定の長さの時系列データを複数作成する時系列データ作成部

(2) 複数の上記時系列データのそれぞれについて、前記指定の長さの複数の時系列データに対して、平均をとる為のセグメント幅を決める中間次元決定部

5 (3) 前記各時系列に対して、上記平均をとる為のセグメント幅ごとに
平均値を計算する平均値計算実行部

(4) 計算した上記平均値をセグメント代表値とした中間時系列を作成する中間時系列作成部

(5) それぞれの中間時系列を対象として特異値分解を行う特異値分解
実行部

10 (6) 指定された数の上位の特異値分解の成分を上記中間時系列の圧縮
データとする次元圧縮時系列データ生成部。

図 1

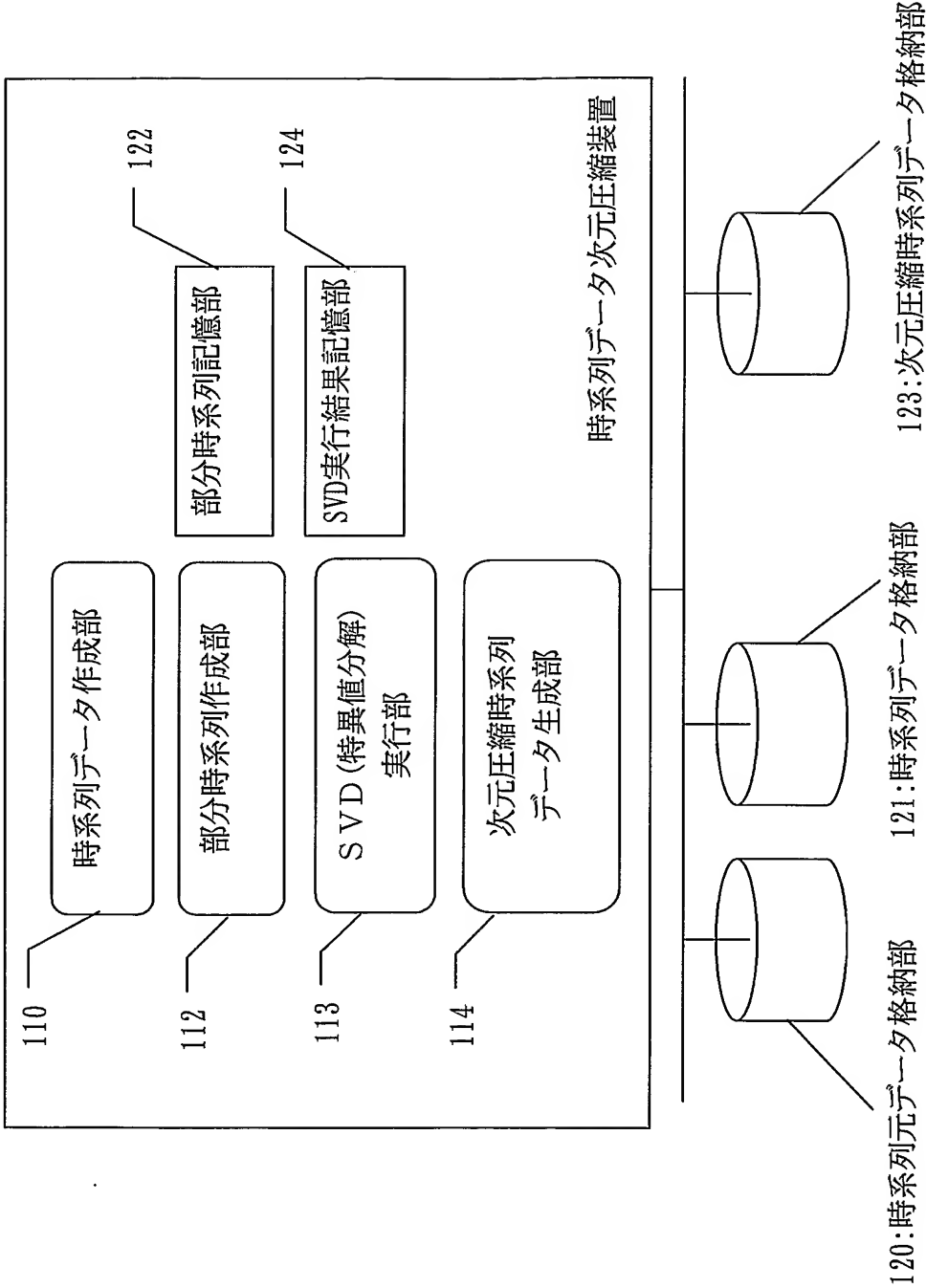


図 2

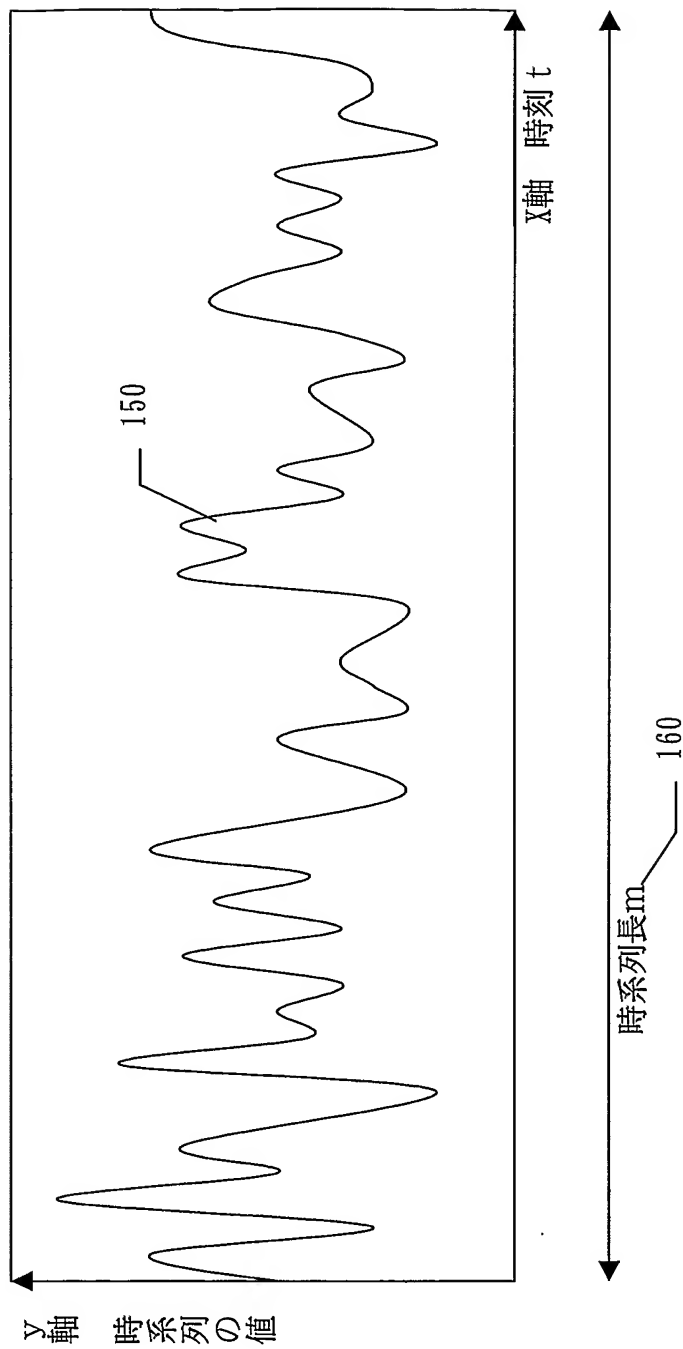


図 3

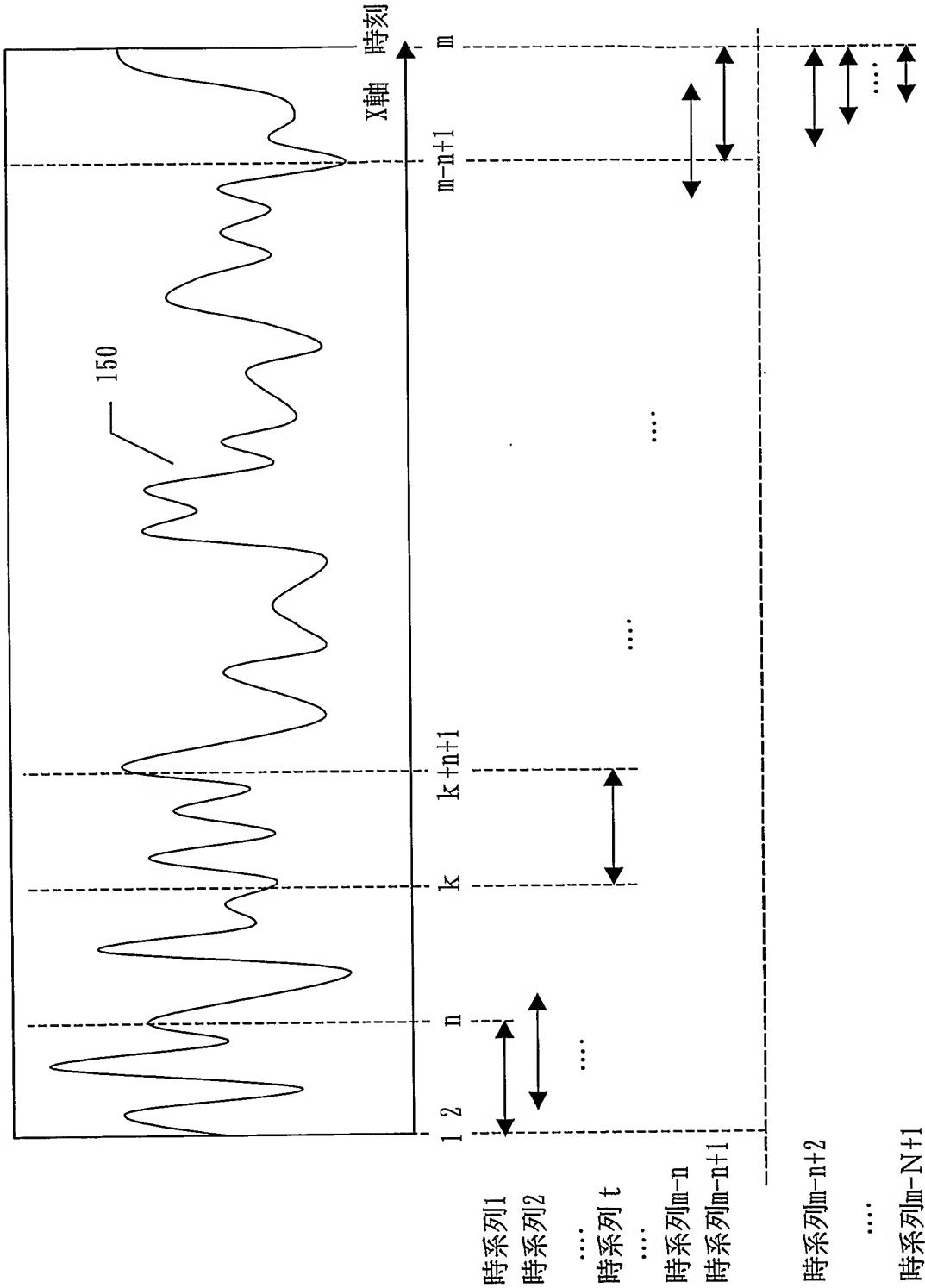


図 4

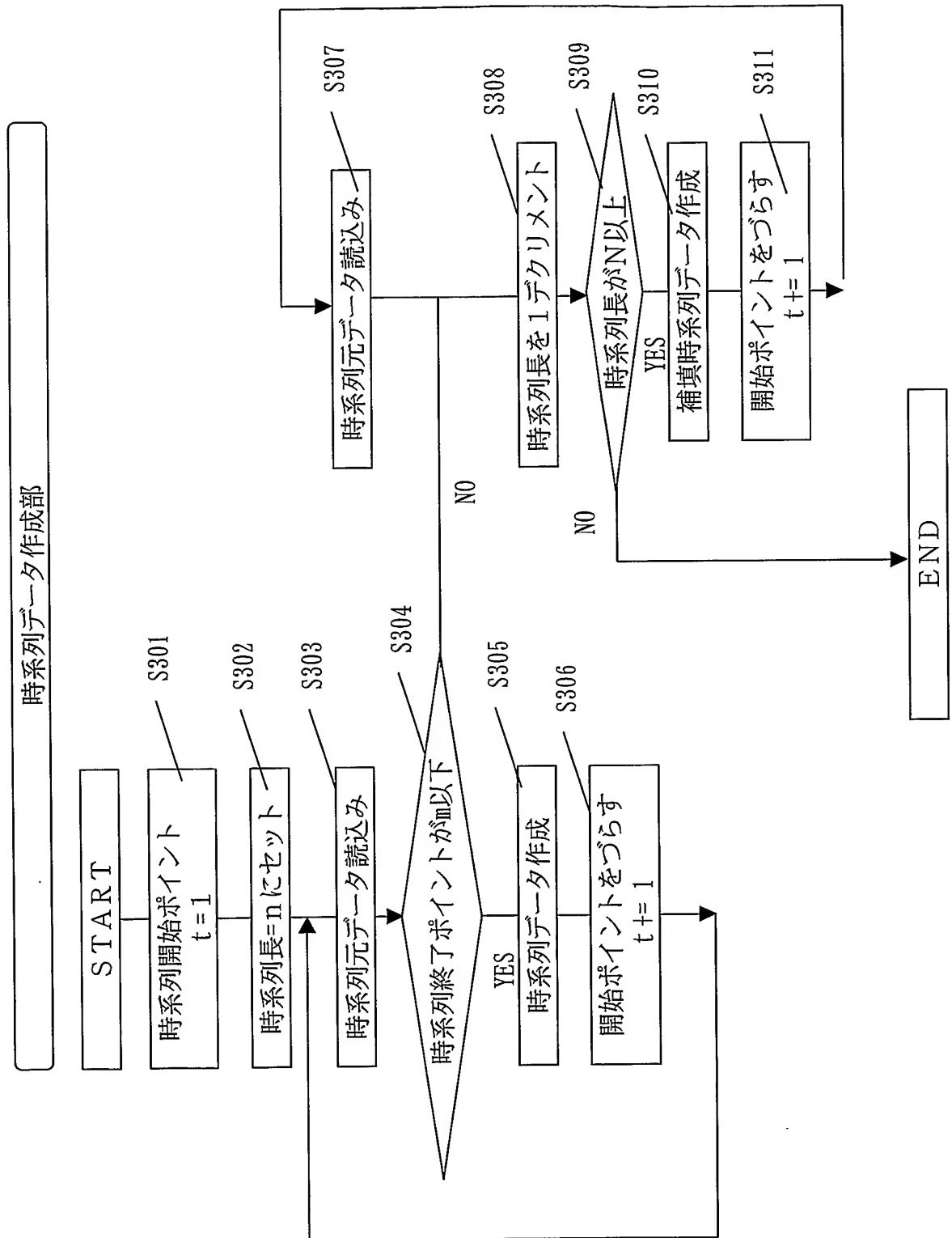


図 5

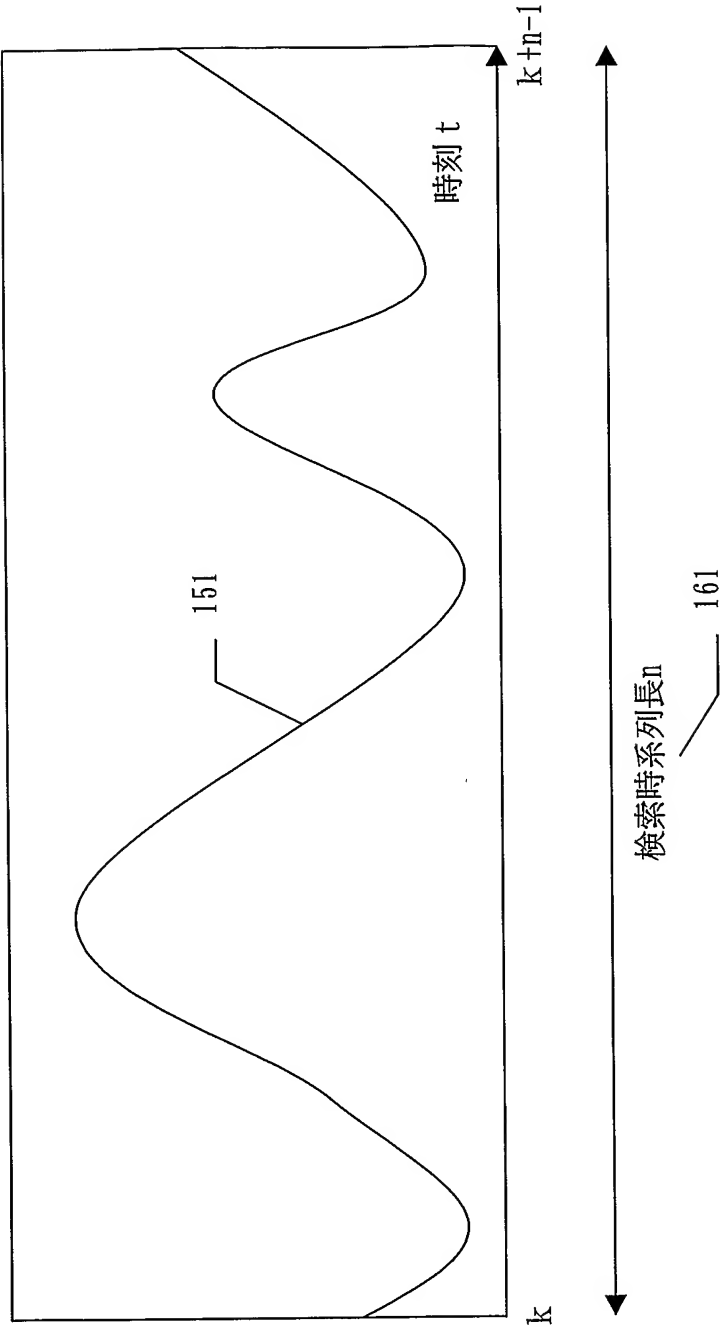


図 6

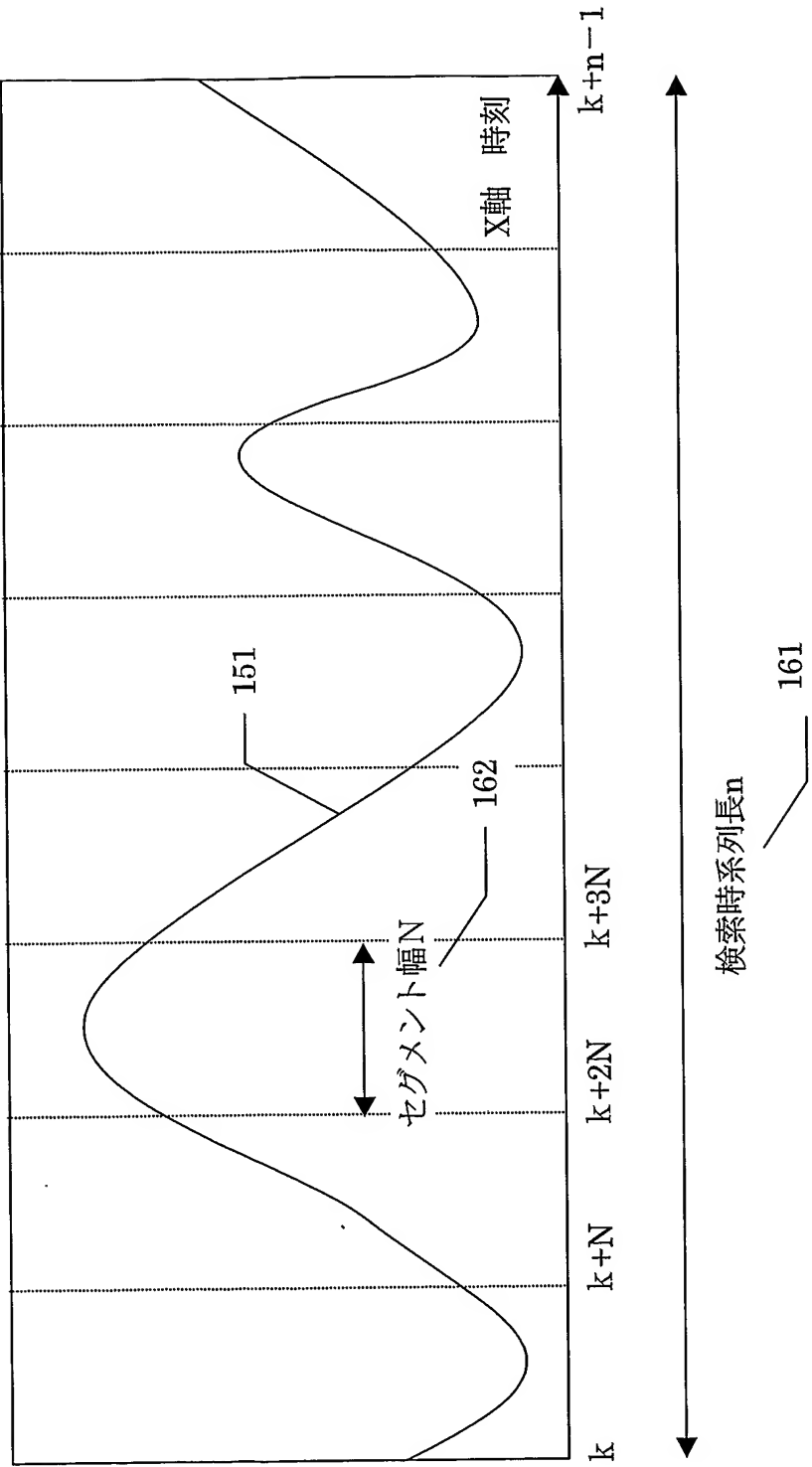


図 7

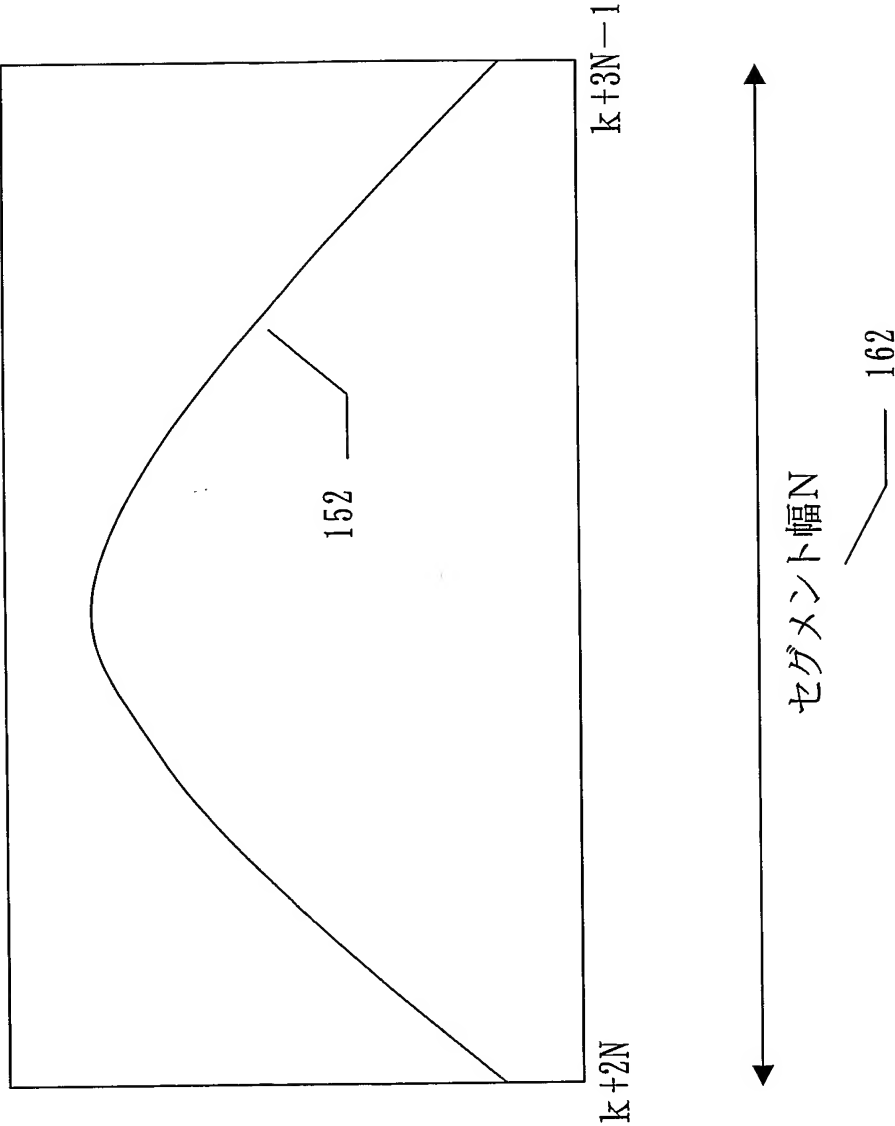


図 8

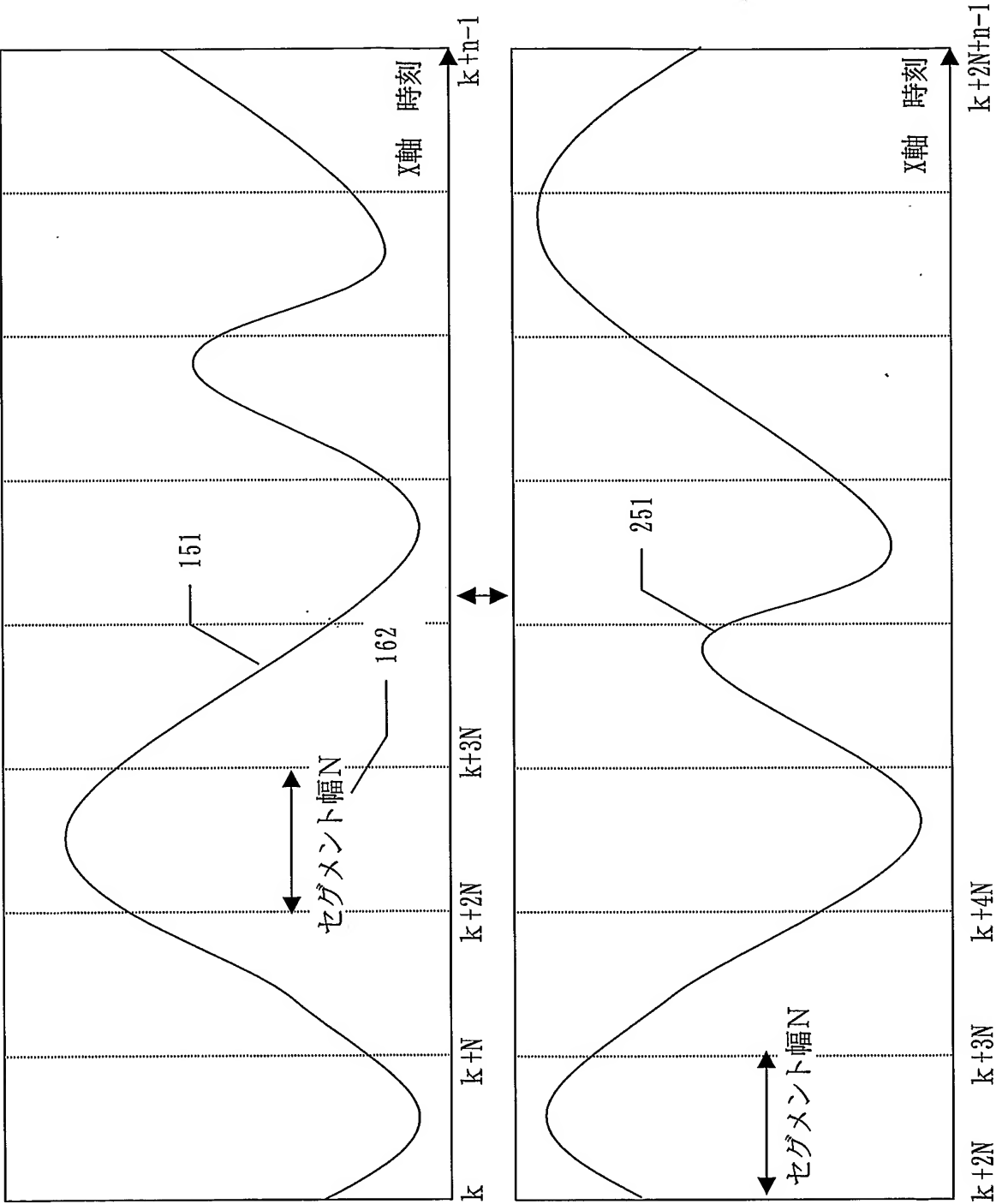
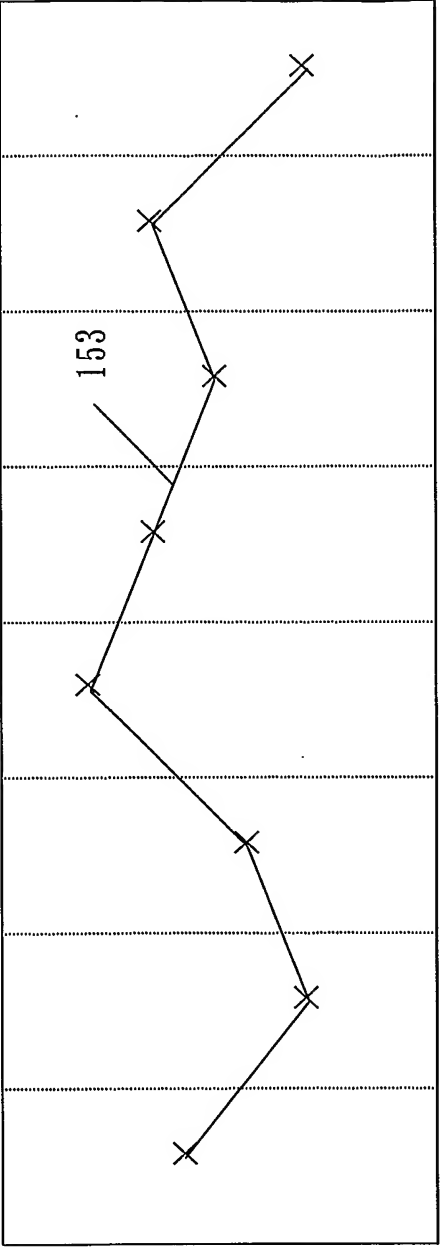


図 9

部分時系列開始 ポイント	固有値 (s_1)	u_1 の要素の値
1	2904.518	0.04495
2	2904.518	0.04497
...
r	2904.518	0.04350
...
m-N	2904.518	0.04371
m-N+1	2904.518	0.04385

図 10

次元圧縮時系列データ



11/20

図 11

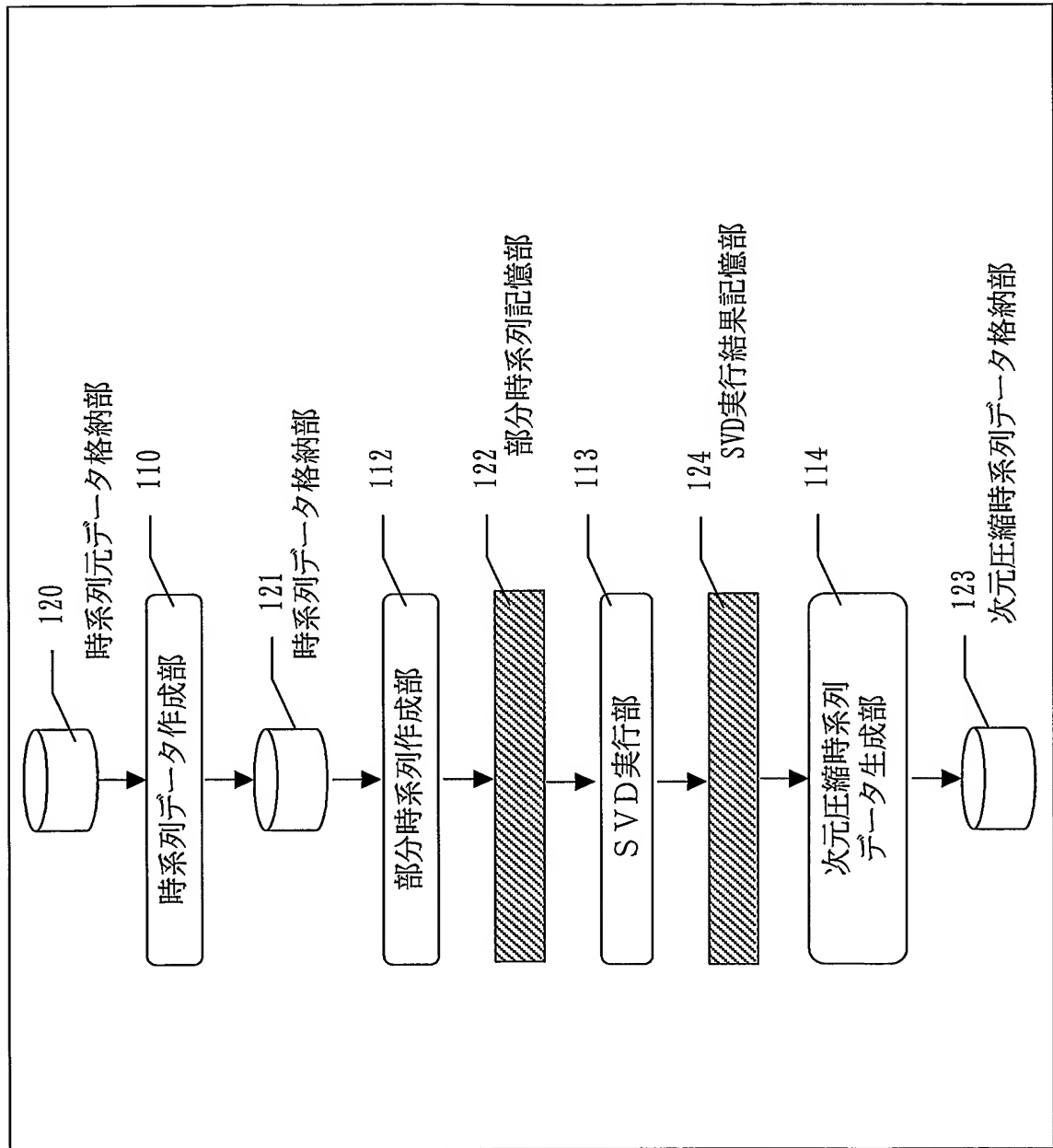


図 12

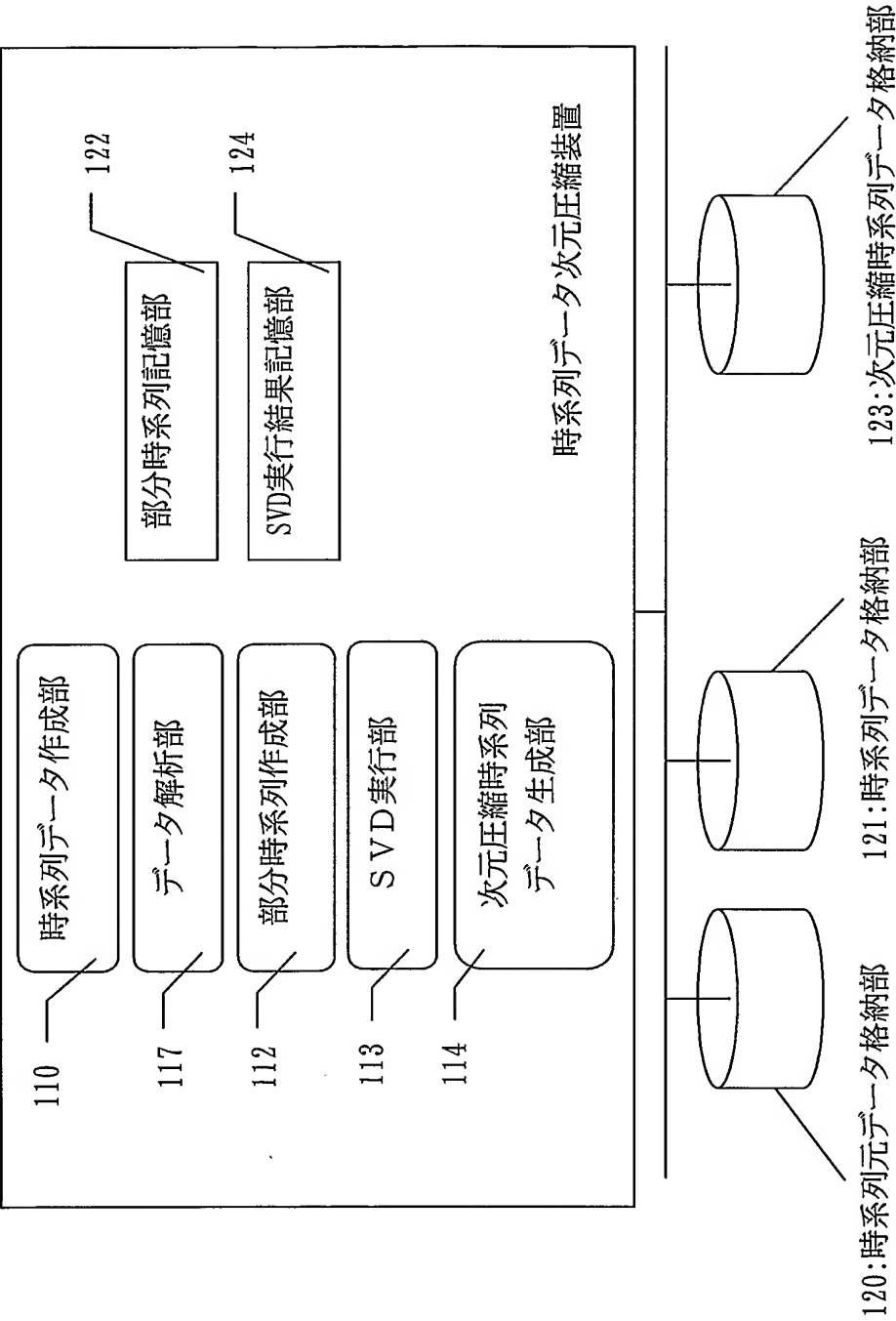


図 13

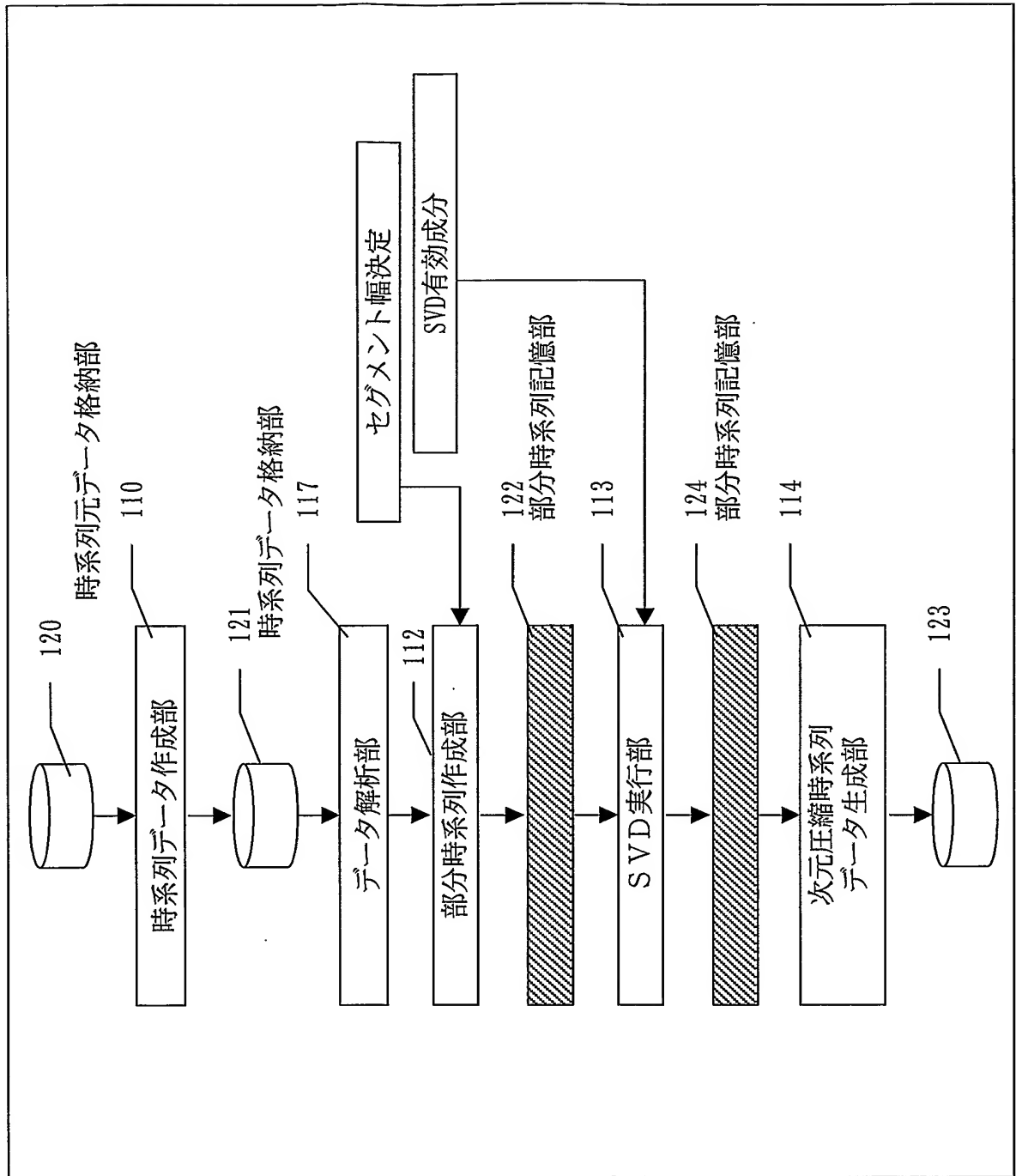


図 14

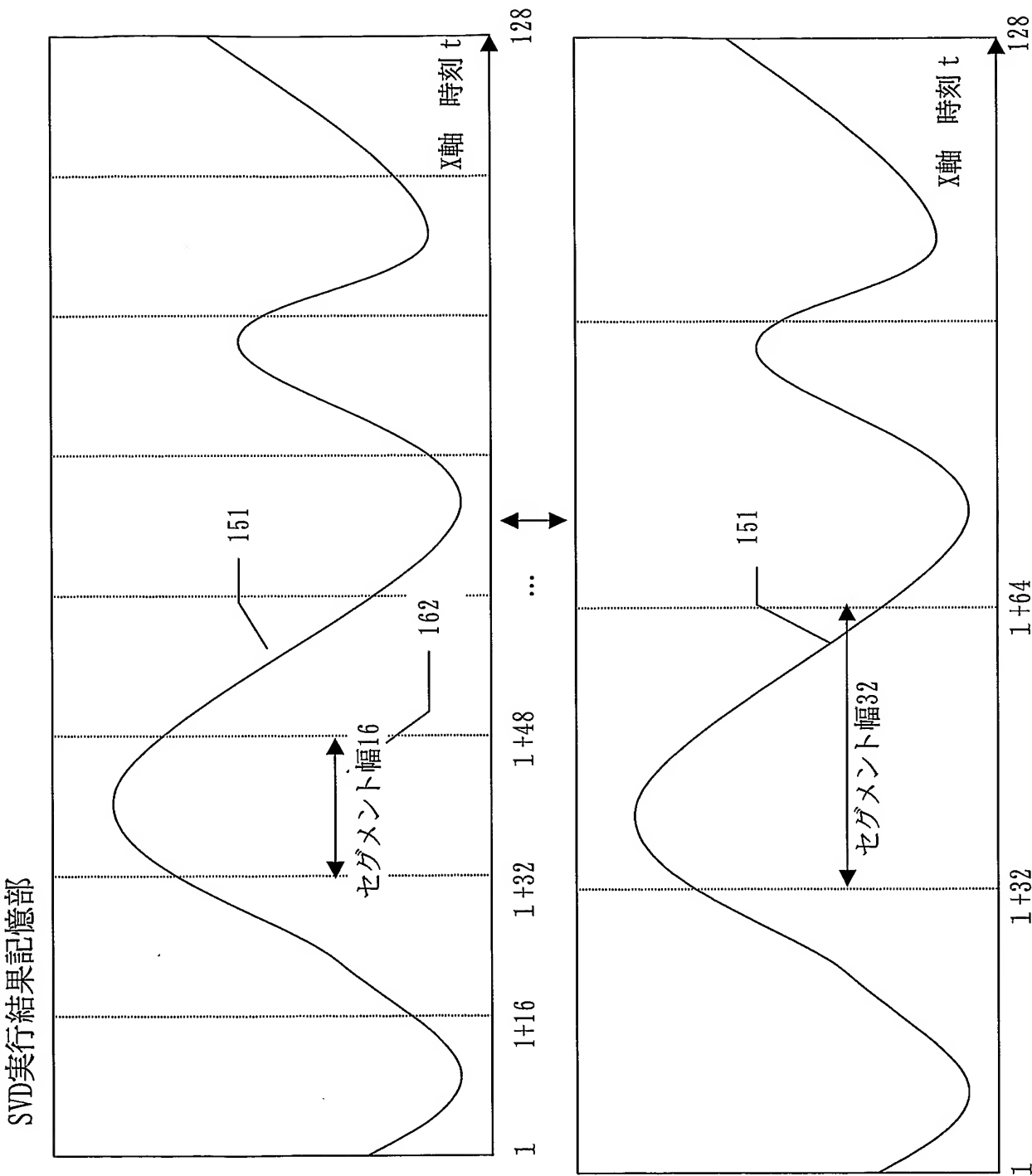


図 15

部分時系列開始ポイント	固有値 (s_1)	u_1 の要素の値	固有値 (s_2)	u_2 の要素の値
1	2904.518	0.04495	458.845	-0.05151
2	2904.518	0.04497	458.845	-0.05616
...
r 1 5 0	2904.518	0.04350	458.845	0.04443
...
m-N	2904.518	0.04371	458.845	0.02856
m-N+1	2904.518	0.04385	211.45	0.01931

図 16

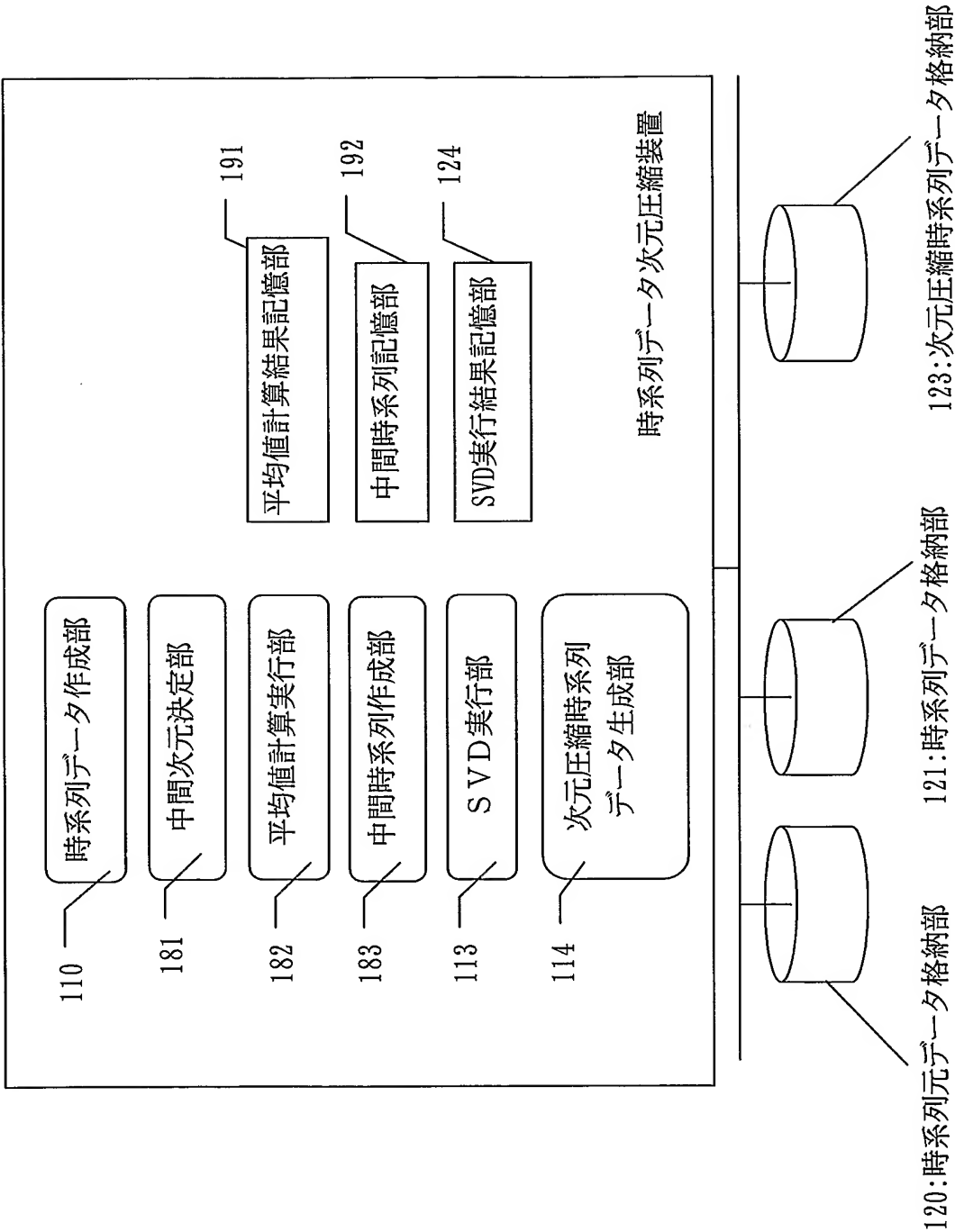
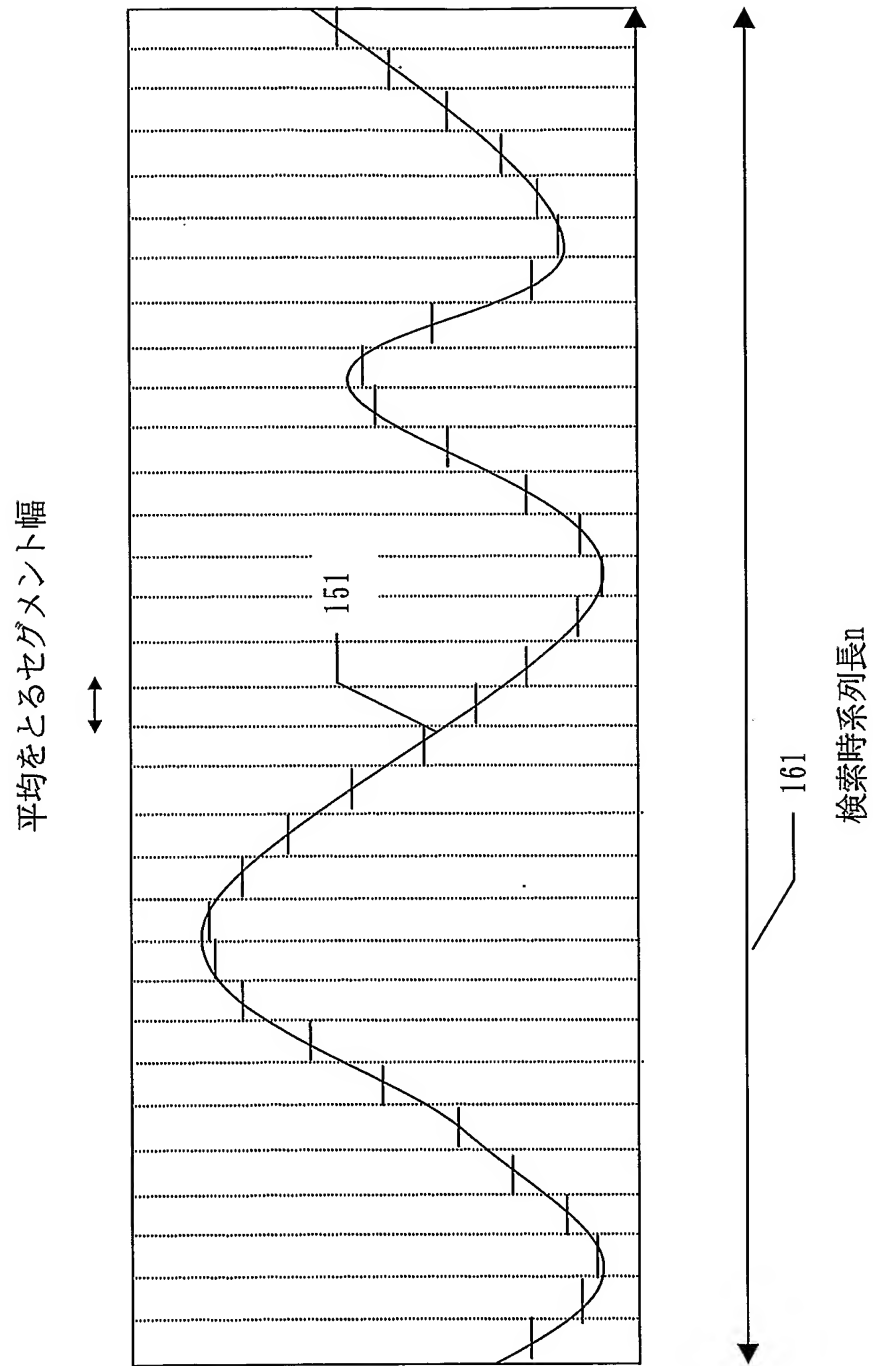


図 17



18/20

図 18

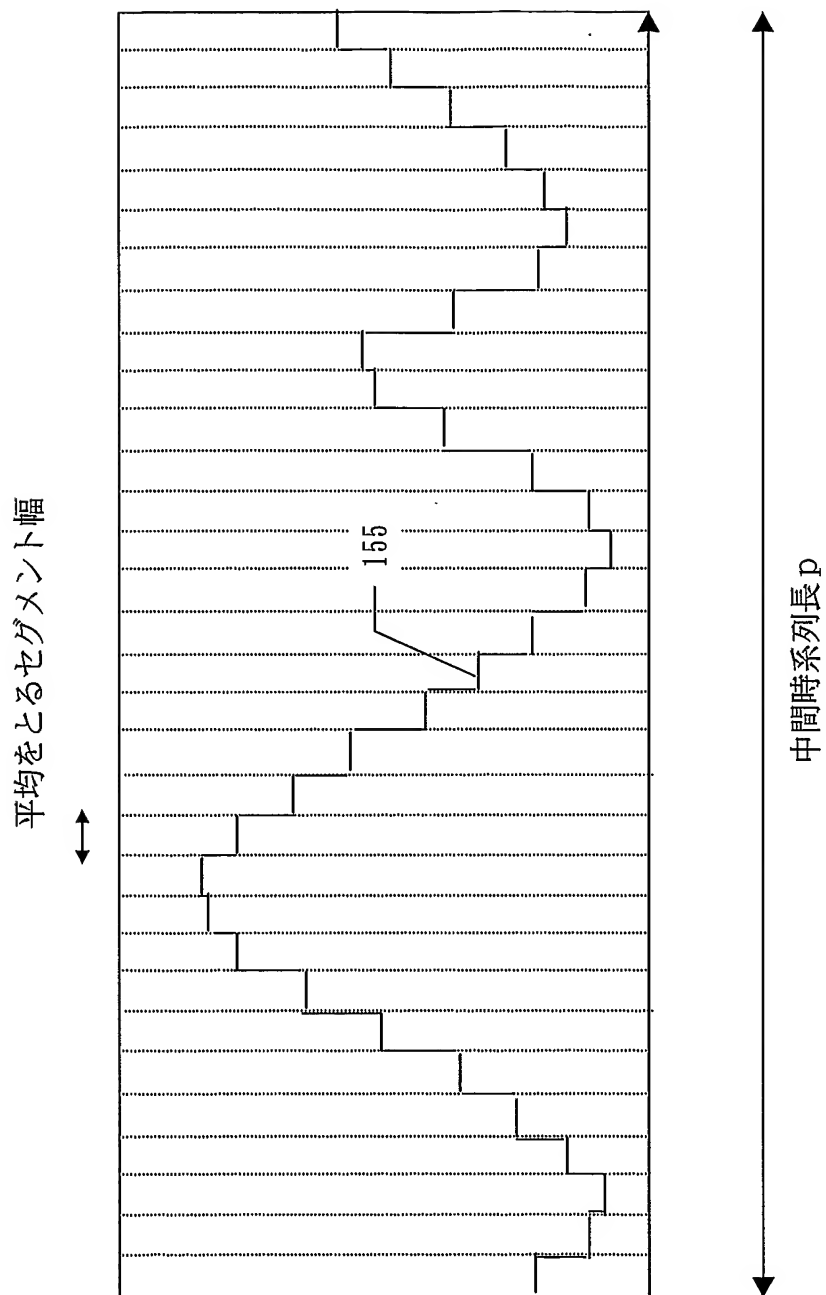
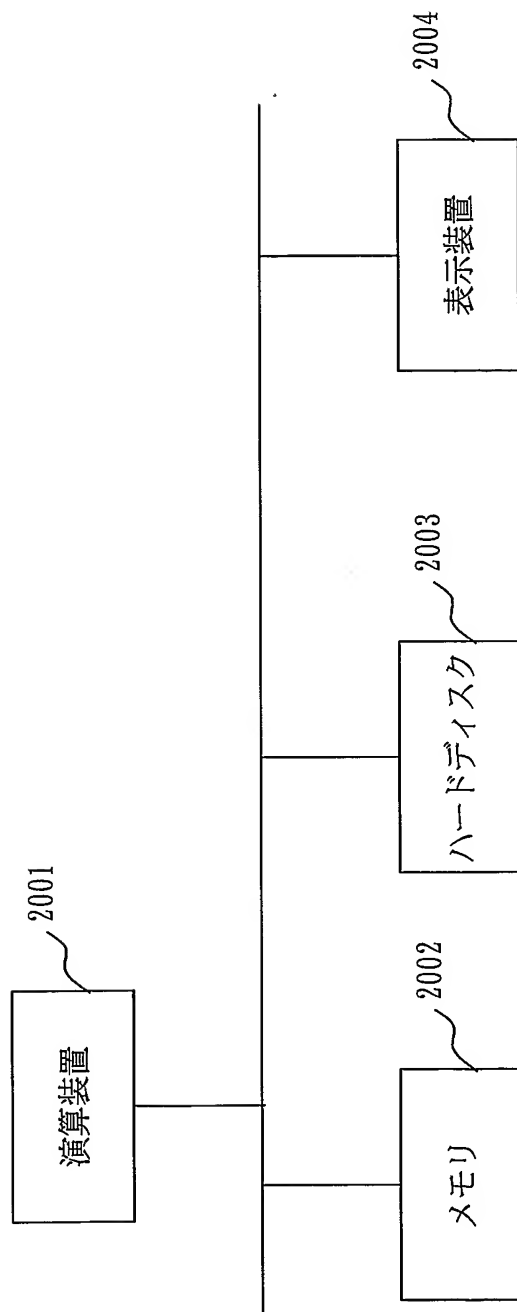


図 19

中間時系列開 始ポイント	固有値 (s_1)	u_1 の要素の 値	固有値 (s_2)	u_2 の要素の 値	...	固有値 (s_8)	u_8 の要素 の値
1	2904.518	0.04495	458.845	-0.05151
2	2904.518	0.04497	458.845	-0.05616
...
150	2904.518	0.04350	458.845	0.04443
...
m·n	2904.518	0.04371	458.845	0.02856
m·n+1	2904.518	0.04385	211.45	0.01931

20/20

図 20



INTERNATIONAL SEARCH REPORT

International application No.

PCT/JP2004/002252

A. CLASSIFICATION OF SUBJECT MATTER

Int.Cl⁷ 0H3M7/30, G06F17/30

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

Int.Cl⁷ 0H3M7/30, G06F17/30

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Jitsuyo Shinan Koho	1926-1996	Toroku Jitsuyo Shinan Koho	1994-2004
Kokai Jitsuyo Shinan Koho	1971-2004	Jitsuyo Shinan Toroku Koho	1996-2004

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	K. Chakrabarti, E. Keogh, S. Mehrotra, M. Pazzani, Locally Adaptive Dimensionality Reductoin for Indexing Large Time Series Data bases, ACM Tran sactions on Database Systems, Vol.27, No.2, June 2002, pages 188 to 228	1-4
A	E. Keogh, K. Chakrabarti, M. Pazzani, S. Mehrotra, Dimensionality Reduction for Fast Similarity Search in Large Time Series Databases, Knowledge and Information Systems, Vol.3, No.3, 2000, pages 263 to 286	1-4
A	F. Korn, H.V., Jagadish, C., Faloutsos, Efficiently Supprting Ad Hoc Queries in Large Datasets of Time Sequences, Proceedings of SIGMOD '97, 1997, pages 289 to 300	1-4



Further documents are listed in the continuation of Box C.



See patent family annex.

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search

08 April, 2004 (08.04.04)

Date of mailing of the international search report

20 April, 2004 (20.04.04)

Name and mailing address of the ISA/
Japanese Patent Office

Authorized officer

Facsimile No.

Telephone No.

INTERNATIONAL SEARCH REPORT

International application No.

PCT/JP2004/002252

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	JP 11-288294 A (Honda Motor Co., Ltd.), 19 October, 1999 (19.10.99), Full text; all drawings (Family: none)	1-4

A. 発明の属する分野の分類 (国際特許分類 (IPC))

Int. Cl⁷ H03M7/30, G06F17/30

B. 調査を行った分野

調査を行った最小限資料 (国際特許分類 (IPC))

Int. Cl⁷ H03M7/30, G06F17/30

最小限資料以外の資料で調査を行った分野に含まれるもの

日本国実用新案公報 1926-1996年
 日本国公開実用新案公報 1971-2004年
 日本国登録実用新案公報 1994-2004年
 日本国実用新案登録公報 1996-2004年

国際調査で使用した電子データベース (データベースの名称、調査に使用した用語)

C. 関連すると認められる文献

引用文献の カテゴリー*	引用文献名 及び一部の箇所が関連するときは、その関連する箇所の表示	関連する 請求の範囲の番号
A	K. Chakrabarti, E. Keogh, S. Mehrotra, M. Pazzani, Locally Adaptive Dimensionality Reduction for Indexing Large Time Series Data bases, ACM Transactions on Database Systems, Vol. 27, No. 2, June 2002, pages 188-228	1-4
A	E. Keogh, K. Chakrabarti, M. Pazzani, S. Mehrotra, Dimensionality Reduction for Fast Similarity Search in Large Time Series Databases, Knowledge and Information Systems, Vol. 3, No. 3, 2000, pages 263-286	1-4

☒ C欄の続きにも文献が列挙されている。☐ パテントファミリーに関する別紙を参照。

* 引用文献のカテゴリー

「A」 特に関連のある文献ではなく、一般的技術水準を示すもの
 「E」 国際出願日前の出願または特許であるが、国際出願日後に公表されたもの
 「L」 優先権主張に疑義を提起する文献又は他の文献の発行日若しくは他の特別な理由を確立するために引用する文献 (理由を付す)
 「O」 口頭による開示、使用、展示等に言及する文献
 「P」 国際出願日前で、かつ優先権の主張の基礎となる出願

の日の後に公表された文献

「T」 国際出願日又は優先日後に公表された文献であって出願と矛盾するものではなく、発明の原理又は理論の理解のために引用するもの
 「X」 特に関連のある文献であって、当該文献のみで発明の新規性又は進歩性がないと考えられるもの
 「Y」 特に関連のある文献であって、当該文献と他の1以上の文献との、当業者にとって自明である組合せによって進歩性がないと考えられるもの
 「&」 同一パテントファミリー文献

国際調査を完了した日

08.04.2004

国際調査報告の発送日

20.4.2004

国際調査機関の名称及びあて先

日本国特許庁 (ISA/JP)
 郵便番号100-8915
 東京都千代田区霞が関三丁目4番3号

特許庁審査官 (権限のある職員)

北村 智彦

5K

9297

電話番号 03-3581-1101 内線 3555

C (続き) . 関連すると認められる文献		
引用文献の カテゴリー*	引用文献名 及び一部の箇所が関連するときは、その関連する箇所の表示	関連する 請求の範囲の番号
A	F. Korn, H. V. Jagadish, C. Faloutsos, Efficiently Supprting Ad Hoc Queries in Larage Datasets of Time Sequences, Proceedings of SIGMOD '97, 1997, pages. 289-300	1-4
A	J P 11-288294 A (本田技研工業株式会社) 1999. 10. 19, 全文, 全図 (ファミリーなし)	1-4